Methods in Ecology and Evolution

British Ecological Society

Methods in Ecology and Evolution 2013 doi: 10.1111/2041-210X.12037

High sensitivity of 454 pyrosequencing for detection of rare species in aquatic communities

Aibin Zhan[1,2]*, Martin Hulák[1], Francisco Sylvester[3], Xiaoting Huang[4], Abisola A. Adebayo[1], Cathryn L. Abbott[5], Sarah J. Adamowicz[6], Daniel D. Heath[1], Melania E. Cristescu[1,7] and Hugh J. MacIsaac[1]

[1]*Great Lakes Institute for Environmental Research, University of Windsor, 401 Sunset Avenue, Windsor, Ontario N9B 3P4, Canada;* [2]*Research Center for Eco-Environmental Sciences, Chinese Academy of Sciences, 18 Shuangqing Road, Haidian District, Beijing 100085, China;* [3]*Facultad de Ciencias Exactas y Naturales, Departamento de Ecología, Genética y Evolución and Consejo Nacional de Investigaciones Científicas y Técnicas, Universidad de Buenos Aires, Buenos Aires, Argentina;* [4]*Key Laboratory of Marine Genetics and Breeding (MGB), Ministry of Education, College of Marine Life Sciences, Ocean University of China, 5 Yushan Road, Qingdao 266003, China;* [5]*Pacific Biological Station, Fisheries and Oceans Canada, 3190 Hammond Bay Road, Nanaimo, British Columbia V9T 6N7, Canada;* [6]*Biodiversity Institute of Ontario and Department of Integrative Biology, University of Guelph, 50 Stone Road East, Guelph, Ontario N1G 2W1, Canada; and* [7]*Department of Biology, McGill University, 1205 Docteur Penfield, Montreal, Quebec H3A 1B1, Canada*

Summary

1. Concerns regarding the rapid loss of endemic biodiversity, and introduction and spread of non-indigenous species, have focused attention on the need and ability to detect species present in communities at low abundance. However, detection of rare species poses immense technical challenges, especially for morphologically cryptic species, microscopic taxa and those beneath the water surface in aquatic ecosystems.

2. Next-generation sequencing technology provides a robust tool to assess biodiversity, especially for detection of rare species. Here, we assess the sensitivity of 454 pyrosequencing for detection of rare species using known indicator species spiked into existing complex plankton samples. In addition, we develop universal small subunit ribosomal DNA primers for amplification of a wide range of taxa for detailed description of biodiversity in complex communities.

3. A universality test of newly designed primers for the hypervariable V4 region of the nuclear small subunit ribosomal DNA (V4-nSSU) using a plankton sample collected from Hamilton Harbor showed that 454 pyrosequencing based on this universal primer pair can recover a wide range of taxa, including animals, plants (algae), fungi, blue-green algae and protists.

4. A sensitivity test showed that 454 pyrosequencing based on newly designed universal V4-nSSU primers was extremely sensitive for detection of very rare species. Pyrosequencing was able to recover spiked indicator species with biomass percentage as low as approximately $2{\cdot}3 \times 10^{-5}\%$ when 24 artificially assembled samples were tagged and sequenced in one PicoTiter plate (i.e. sequencing depth of an equivalent of 1/24 PicoTiter plate). In addition, spiked rare species were sometimes recovered as singletons (i.e. Operational Taxonomic Units represented by a single sequence), suggesting that at least some singletons are informative for recovering unique lineages in 'rare biospheres'.

5. The method established here allows biologists to better investigate the composition of aquatic communities, especially for detection of rare taxa. Despite a small-scale pyrosequencing effort, we demonstrate the extreme sensitivity of pyrosequencing using rare species spiked into plankton samples. We propose that the method is a powerful tool for detection of rare native and/or alien species.

Key-words: alien invasive species, biodiversity, endangered species, next-generation sequencing (NGS), non-indigenous species, nSSU rDNA, plankton, universal primers

Introduction

Global biodiversity has been greatly impacted over the past century by a range of interacting stressors, including climate change, over-exploitation, habitat loss, chemical pollution and introductions of non-indigenous, invasive species (see review by Pereira, Navarro & Martins 2012). These profound changes have intensified the need to investigate causes and consequences of rapid changes of community structure and composition. Such an understanding is predicated on our ability to accurately determine the

*Correspondence author. E-mail: zhanaibin@hotmail.com, azhan@rcees.ac.cn

entire complement of species present before and after stressors have begun to impact communities.

One of the most important aspects of biodiversity conservation is to manage rare species in communities (Lodge *et al.* 2006; Darling & Mahon 2011; Jerde *et al.* 2011). Two types of species may occur in the 'rare biosphere' which is constituted by diverse, low-abundance species (Sogin *et al.* 2006): native rare species, which may (or may not) be endangered, and recently introduced non-indigenous species (NIS). Rare species may be vulnerable to extirpation owing to demographic stochasticity or ecosystem degradation (Wilson *et al.* 2011). On the other hand, establishment of invasive NIS may accompany or contribute to ecosystem degradation and drive native species extinct (e.g. Ricciardi 2007; Pyšek & Richardson 2010). Recently introduced NIS are typically present at very low population size, even in cases where they eventually become dominant (see Crooks & Soulé 1999 and examples therein). Consequently, effective conservation plans must identify the complement of native rare species in need of protection, as well as recently established NIS that may adversely affect native biodiversity.

The detection of rare species represents technical challenges in all environments, but particularly so in aquatic ecosystems where populations may be small, geographically constrained and hidden beneath the water surface (McDonald 2004; Jerde *et al.* 2011). Traditional methods rely on capture using different sampling instruments and then identification *via* assessment of morphological and anatomical characteristics (Gu & Swihart 2004). These traditional methods can only consistently detect species at moderate-to-high abundance in communities, even for large animals (Magnuson, Benson & McLain 1994). For example, Harvey, Qureshi & MacIsaac (2009) found that detection rate could be <2% for newly introduced NIS using bulk samples, but was dependent on both sampling effort and population density. Many NIS remain inconspicuous at subthreshold densities until they experience population growth and are detected (Harvey, Qureshi & MacIsaac 2009; Jerde *et al.* 2011). In addition, correct identification is often difficult or impossible, especially for cryptic species complexes, species with sexual dimorphism, and for particular life stages such as eggs and immature individuals (e.g. Hebert *et al.* 2004; Barrett & Hebert 2005; Briski *et al.* 2011; Mahon *et al.* 2011). Low detection probability may result in type II errors (i.e. false negatives), which is a critical problem when addressing the fate of endangered species, as well as for managers trying to determine if a known NIS has colonized a particular region.

To improve efficiency and accuracy, various DNA-based methods have been successfully developed and applied (see reviews by Medlin & Kooistra 2010; Darling & Mahon 2011; Lodge *et al.* 2012). The efficacy of these DNA-based methods has been demonstrated in detection of both native endangered species and newly introduced NIS (e.g. Mackie & Geller 2010; Jerde *et al.* 2011). The recent advent of second-generation sequencing technology, such as 454 pyrosequencing, has radically changed our ability to identify biodiversity in communities (e.g. Creer 2010; Fonseca *et al.* 2010). Advantages including extremely deep sequencing, low cost per sequence read and high throughput render this technology a promising tool for deep investigation of biodiversity in communities, especially in communities that contain rare, inconspicuous species.

Plankton biodiversity, especially the 'rare biosphere', remains poorly described and underestimated (e.g. Fonseca *et al.* 2010). Next-generation sequencing technology, especially 454 pyrosequencing, has been applied to identify species composition of soil microbes (Shade *et al.* 2012), aquatic metazoans (Fonseca *et al.* 2010; Shokralla *et al.* 2011), human gut flora (Yatsunenko *et al.* 2012), and other communities. However, this newly emerged technology requires better characterization and efficiency testing before it can be broadly applied to biodiversity assessments of plankton. In addition, the lack of universal primers for amplification of a wide range of taxa has hampered detailed description of biodiversity in complex communities. Here, we test the sensitivity of 454 pyrosequencing for detection of rare species using spiked additions of indicator species to existing freshwater and marine plankton samples. Also, we develop new universal primers that may be used in future biodiversity studies for a variety of groups owing to their broad applicability.

Materials and methods

To test the sensitivity of 454 pyrosequencing for detection of rare species, we added known species to existing complex plankton samples using different dilution gradients. These artificially assembled plankton samples were subjected to pyrosequencing to determine whether or not these spiked rare species could be successfully recovered. In addition, we designed a universal primer pair for the hypervariable V4 region of the nuclear small subunit ribosomal DNA (V4-nSSU) for biodiversity assessment based on 454 pyrosequencing. The sensitivity and universality of this primer pair were tested using a small-scale pyrosequencing of a freshwater plankton sample collected from Hamilton Harbor in Lake Ontario.

UNIVERSAL PRIMER DESIGN

For universal primer design for V4-nSSU, we recovered sequences from GenBank (http://www.ncbi.nlm.nih.gov/nuccore) of representative species of the three major groups of interest (Crustacea, Mollusca, Tunicata) owing to their history of invasiveness. In total, we included 142 species to cover almost all orders/suborders of these groups. All downloaded sequences were aligned using MEGA version 5 (Tamura *et al.* 2011), inspected manually, and universal primers were designed in conserved regions (Fig. 2). Based on the read length (~ 500 bp) of the 454 GS-FLX Titanium platform, all primers were designed to amplify approximately 400–600 bp depending on variable length in different species to get maximum information for species identification. The forward primer used for pyrosequencing was tagged specifically for each sample using eight nucleotides to identify pooled PCR products after pyrosequencing (Parameswaran *et al.* 2007). In addition, the 454 FLX adaptors (adaptor A: GCCTCCCTCGCGCCATCAG, adaptor B: GCCTTGCCAGCCCGCTCAG) were also added to the 5′-end of the forward and reverse primers, respectively, to make them compatible with pyrosequencing procedures.

We performed three steps of test for amplification capacity of the universal primers designed for V4-nSSU. First, we tested the universality of primers using several species from each taxonomic

group studied: representative members of crustaceans included *Daphnia pulex*, *Cercopagis pengoi* and *Carcinus maenas*, while molluscs were represented by *Limnoperna fortunei* and *Dreissena polymorpha*, and tunicates by *Ciona intestinalis* and *Botrylloides violaceus*. Sequences from these species had not been included in the alignment used for primer design. Secondly, the primers that performed well in the first step were then tested on bulk DNA isolated from a plankton sample collected from Hamilton Harbor in Lake Ontario. The resulting PCR products were cloned into a vector using a TA cloning kit (Invitrogen Inc., ON, Canada). Twenty-four clones were randomly selected and sequenced using traditional Sanger sequencing method to verify whether or not the selected primers could amplify multiple species when presented simultaneously in a plankton sample. Finally, we employed a small-scale run of 454 pyrosequencing (i.e. an equivalent of 1/48 PicoTiter plate) to assess the performance of the selected primers for biodiversity assessment using the same bulk DNA as that used for Sanger sequencing.

BIOLOGICAL SAMPLE PREPARATION, PYROSEQUENCING AND DATA ANALYSIS

We spiked larvae/juveniles of four species, including two marine species (bay scallop *Argopecten irradians* and Japanese sea cucumber *Apostichopus japonicus*) and two freshwater species (golden mussel *Limnoperna fortunei* and water lice *Asellus aquaticus*), into plankton samples to test the sensitivity of 454 pyrosequencing for detection of rare species. To avoid possible errors and confusion derived from spiked species, we spiked marine species into freshwater plankton samples and freshwater species into marine plankton samples.

The plankton samples were collected from major ports in the Great Lakes (Hamilton, Nanticoke and Thunder Bay) and on the Atlantic coast of North America (Bayside and Hawksbury). We used geo-referenced 80 μm oblique plankton nets to tow from the bottom to water surface in each port to collect plankton samples. Larvae of the two marine species were artificially cultured in the laboratory following Zhan *et al*. (2008), while the golden mussel and water lice were collected from the wild in South America and Europe respectively. All larvae/juveniles were taxonomically confirmed and measured under a microscope. The size of larvae/juveniles of these four species varies from approximately 70 μm to 2 000 μm, and weight, which was averaged based on multiple individuals weighted, ranges from $1{\cdot}8 \times 10^{-4}$ mg to 2 mg (Table 1). All collected samples were immediately preserved in 95% ethanol.

Tubes containing preserved plankton were centrifuged at 12 000 rpm for 3 min to remove ethanol, and then opened in a fume hood for 10–15 min to evaporate residual ethanol. Depending on the available amount of plankton from each port, 50–150 mg of plankton sample was used for DNA isolation. We ran three replicates and four gradients for each spiked species to assess the recovery performance (Fig. 1). For the three smaller species (i.e. bay scallop, Japanese sea cucumber, golden mussel), we established a gradient of 0·01, 0·1, 1 and 5 larva(e) per plankton sample, while the larger water lice was added at 0·001, 0·01, 0·05 and 0·1 individual per plankton sample. All artificial assembling procedures were performed before DNA extraction. For the gradients using ≥ 1 larvae, we spiked larvae directly into plankton samples, while for those <1, we lysed one larva/juvenile using 200 μL DNA lysis buffer and then added different amount of lysed larva/juvenile solution into corresponding lysed plankton samples based on dilution gradients (Fig. 1).

We extracted total genomic DNA using DNeasy Blood and Tissue Kit (Qiagen Inc., ON, Canada). The quality and quantity of each DNA sample were measured by a NanoDrop spectrophotometer (NanoDrop Technologies, DE, USA). We prepared PCR mixtures (25 μL) in eight duplicates for each sample to avoid biased amplification. Each duplicate consisted of 100 ng of genomic DNA, 1 × PCR buffer, 2 mM of Mg^{2+}, 0·2 mM of dNTPs, 0·4 μM of each primer, and 2 U of *Taq* DNA polymerase (Genscript). PCR cycling parameters consisted of an initial denaturation step at 95 °C for 5 min, followed by 25 amplification cycles of 95 °C for 30 s, 50 °C for 30 s, 72 °C for 90 s, and a final elongation step at 72 °C for 10 min. We pooled and purified PCR products of duplicates using the Solid Phase Reversible Immobilization (SPRI) paramagnetic bead-based method (Agencourt Bioscience Corporation, MA, USA).

For pyrosequencing, we pooled PCR products derived from 24 artificially assembled communities to form one PicoTiter plate (totally two plates for 48 assembled samples, Fig. 1). To ensure approximately equal contributions from each sample, equimolar PCR products from each sample were pooled together. Pyrosequencing was performed using 454 FLX Adaptor A on a GS-FLX Titanium platform (454 Life Sciences, CT, USA) by Engencore at the University of South Carolina.

After pyrosequencing, each sample was sorted based on its unique tag labelled on the forward primer using software CLOTU (Kumar *et al*. 2011). Raw sequence reads were denoised, trimmed and filtered prior to subsequent analyses to eliminate errors/artefacts using both RDP pyrosequencing pipeline (http://rdp.cme.msu.edu/) and software CLOTU. In general, we deleted sequence reads with Phred quality scores <20 (Q20), and then we removed sequences that: (i) did not perfectly match the tags and forward primer (10–16% of reads removed); (ii) contained any undetermined nucleotide (N's, 4–6% of reads removed); and (iii) were too short (<250 bp, 4–15% of reads removed). In addition, given that PCR-mediated recombination in amplification products (i.e. chimeras) can inflate species diversity, we identified and then deleted chimeras from each data set (23–35% of OTUs removed) using the newly developed, fast and sensitive algorithm UCHIME (Edgar *et al*. 2011).

Sequence reads from each sample were clustered into similarity-based OTUs at a range of genetic divergence from 1% to 10% (insertions and deletions included) using the CD-HIT method (Li & Godzik 2006) implemented in software CLOTU. The CD-HIT method is based on a heuristic search strategy and offers the capacity for rapid clustering of large similar sequence data sets. OTUs were grouped taxonomically (by suborder or higher, such as Copepoda, Cladocera, etc.) by searching against the nucleotide database of GenBank using MEGABLAST with the parameters of E value $< 10^{-50}$ and minimum query coverage >80%. Spiked rare species were also identified by MEGABLAST from each dilution gradient and replicate using available reference sequences.

Results

UNIVERSALITY OF V4-NSSU PRIMERS

Our alignment of available nSSU sequences from GenBank revealed conserved regions for universal primer design (Fig. 2). We designed one primer pair (Uni18S: AGGGCAAKYCTGGTGCCAGC; Uni18SR: GRCGGTATCTRATCGYCTT) spanning the most polymorphic region of V4 of nSSU. The first primer test step using a broad taxonomic range of species revealed that the newly designed primer pair amplified all species effectively. The second test step using bulk DNA showed that, after cloning and Sanger sequencing,

Table 1. Results for the sensitivity of 454 pyrosequencing for detection of rare species using spiked known indicator species to existing plankton samples. The sequencing depth is an equivalent of 1/24 PicoTiter plate per each artificially assembled sample. Four species, including two marine species (bay scallop *Argopecten irradians*, sea cucumber *Apostichopus japonicus*) and two freshwater species (golden mussel *Limnoperna fortunei*, water lice *Asellus aquaticus*), were spiked into plankton samples collected from major ports in Canada. For each species, three replicates and four gradients were setup to assess the recovery performance. For each replicate, the amount of plankton (in mg) is shown. For successful recovery setups, the amount of plankton samples is bolded, and the percentage of biomass of spiked indicator species and the actual number of sequence reads recovered for the target species are shown

Small species					Large species		
No. of larvae/ juveniles	Replicate no.	Bay scallop *Argopecten irradians* Size: 73·4 ± 2·0 μm Weight: $1{\cdot}8 \times 10^{-4}$ mg	Golden mussel *Limnoperna fortunei* Size: 150·0 ± 20·0 μm Weight: $4{\cdot}0 \times 10^{-4}$ mg	Sea cucumber *Apostichopus japonicus* Size: 1175·4 ± 159·0 μm Weight: $4{\cdot}0 \times 10^{-2}$ mg	No. of larvae/ juveniles	Replicate no.	water lice *Asellus aquaticus* Size: 2020 × 528 μm Weight: 2 mg
0·01	1	100 mg	100 mg	**66 mg ($6{\cdot}1 \times 10^{-4}$%) (3 sequence reads)**	0·001	1	118 mg
	2	74 mg	105 mg	79 mg		2	132 mg
	3	74 mg	107 mg	**76 mg ($5{\cdot}3 \times 10^{-4}$%) (1 sequence read)**		3	**141 mg ($1{\cdot}4 \times 10^{-3}$%) (1 sequence read)**
0·1	1	66 mg	97 mg	**87 mg ($4{\cdot}6 \times 10^{-3}$%) (83 sequence reads)**	0·01	1	112 mg
	2	**78 mg ($2{\cdot}3 \times 10^{-5}$%) (1 sequence read)**	93 mg	**79 mg ($5{\cdot}1 \times 10^{-3}$%) (19 sequence reads)**		2	137 mg
	3	90 mg	96 mg	**83 mg ($4{\cdot}8 \times 10^{-3}$%) (81 sequence reads)**		3	**95 mg ($2{\cdot}1 \times 10^{-2}$%) (1 sequence read)**
1	1	**67 mg ($2{\cdot}6 \times 10^{-4}$%) (1 sequence read)**	**85 mg ($4{\cdot}7 \times 10^{-4}$%) (4 sequence reads)**	**88 mg ($4{\cdot}8 \times 10^{-2}$%) (8 sequence reads)**	0·05	1	**150 mg ($6{\cdot}7 \times 10^{-2}$%) (409 sequence reads)**
	2	95 mg	**68 mg ($5{\cdot}9 \times 10^{-4}$%) (4 sequence reads)**	**74 mg ($5{\cdot}4 \times 10^{-3}$%) (5 sequence reads)**		2	**109 mg ($9{\cdot}1 \times 10^{-2}$%) (37 sequence reads)**
	3	**77 mg ($2{\cdot}3 \times 10^{-4}$%) (1 sequence read)**	**68 mg ($5{\cdot}9 \times 10^{-4}$%) (15 sequence reads)**	**31 mg ($1{\cdot}3 \times 10^{-2}$%) (1521 sequence reads)**		3	**111 mg ($9{\cdot}0 \times 10^{-2}$%) (32 sequence reads)**
5	1	**53 mg ($1{\cdot}7 \times 10^{-3}$%) (3 sequence reads)**	**84 mg ($2{\cdot}4 \times 10^{-3}$%) (68 sequence reads)**	**87 mg (0·23%) (51 sequence reads)**	0·1	1	**110 mg (0·18%) (32 sequence reads)**
	2	**60 mg ($1{\cdot}5 \times 10^{-3}$%) (2 sequence reads)**	**78 mg ($2{\cdot}6 \times 10^{-3}$%) (61 sequence reads)**	**79 mg (0·25%) (84 sequence reads)**		2	**109 mg (0·18%) (42 sequence reads)**
	3	**59 mg ($1{\cdot}5 \times 10^{-3}$%) (5 sequence reads)**	**83 mg ($2{\cdot}4 \times 10^{-3}$%) (58 sequence reads)**	**83 mg (0·24%) (156 sequence reads)**		3	**125 mg (0·16%) (895 sequence reads)**

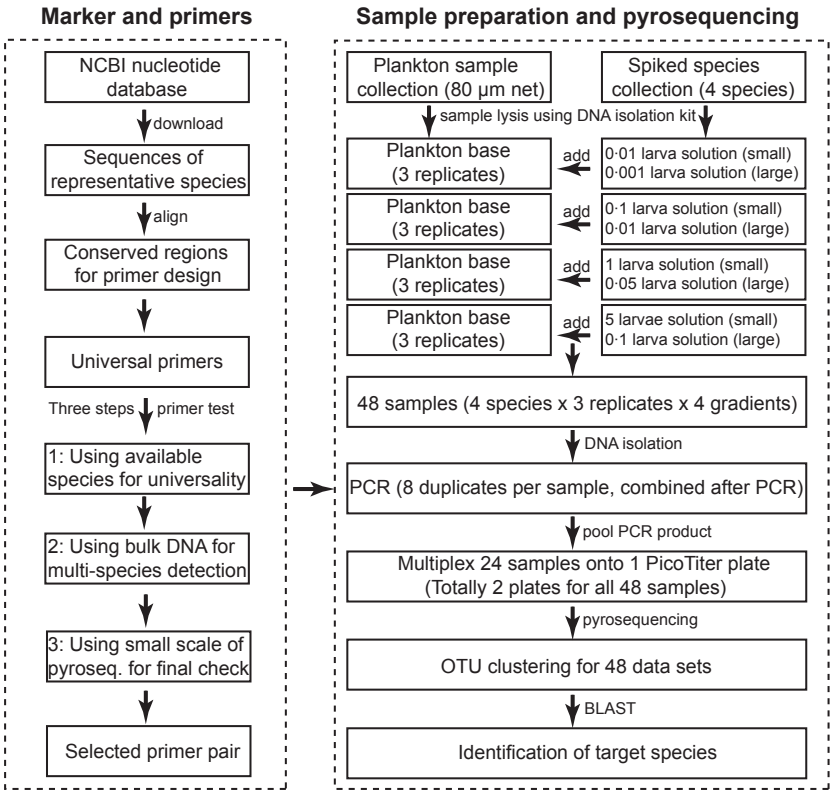

Fig. 1. Methodological flow chart used to design and test universal primers and use them to test the sensitivity of 454 pyrosequencing for detection of rare species. Rare known indicator species were spiked into complex plankton samples.

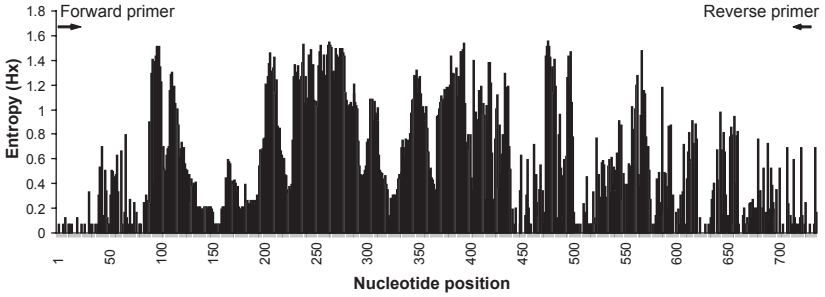

Fig. 2. Information content at each nucleotide position (alignment gaps included) of the hypervariable V4 region of the nuclear small subunit ribosomal DNA (V4-nSSU). Plots are based on aligned sequences of Crustacea, Mollusca and Tunicata deposited in GenBank. Entropy (Hx), which is low in conserved sites and high in variable sites, is shown on the *y*-axis. Universal PCR primers were designed in regions of low entropy (i.e. conserved regions).

the selected primer pair could amplify multiple species of different taxa in the same PCR reaction.

The third primer testing step based on pyrosequencing yielded approximately 15000 sequence reads. After denoising and error/artefact removal, we clustered a total of 6900 high-quality sequence reads (Appendix S1) into similarity-based OTUs using genetic divergences ranging from 1% to 10% (Fig. 3a). Despite a small-scale pyrosequencing run, we detected a high level of biodiversity in the plankton sample, even when high divergence values were used to cluster sequences into OTUs (Fig. 3a, Appendix S1). When these OTUs were assigned to taxonomic groups using BLAST searches, a wide range of taxa was recovered, including animals, plants (algae), fungi, prokaryotes (Cyanobacteria) and protists (data shown at 5% divergence, see Fig. 3b and Appendix S1).

SENSITIVITY OF 454 PYROSEQUENCING FOR DETECTION OF SPIKED RARE SPECIES

Approximately 23 000–25 000 sequence reads were available after error/artefact removal for each sample. BLAST searches revealed that larvae of the smallest animal, the bay scallop ($73{\cdot}4 \pm 2{\cdot}0$ µm), were recovered in all three replicates when the density was five larvae per plankton sample (biomass percentages: $1{\cdot}5–1{\cdot}7 \times 10^{-3}$%, Table 1). It was also detected

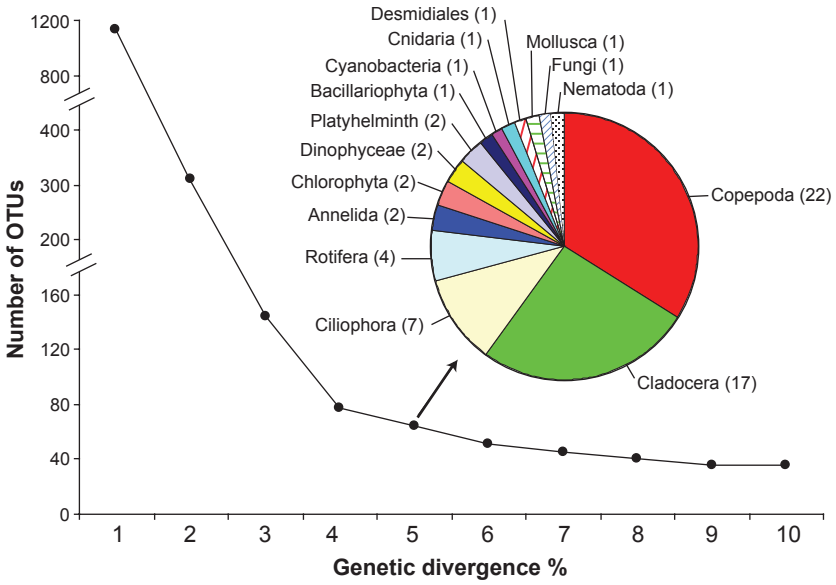

Fig. 3. Number of operational taxonomic units (OTUs) at a range of genetic divergence from 1% to 10% based on a small scale of pyrosequencing run (i.e. an equivalent of 1/48 PicoTiter plate) of a plankton sample collected from Hamilton Harbor. The analysis used the hypervariable V4 region of the small subunit ribosomal DNA (V4-nSSU) (a), and assessed community composition of the plankton sample at 5% genetic divergence (b). The number of OTUs in each taxonomic group is included in brackets after taxonomic names.

in one and two replicates at 0·1 (biomass percentage: $2{\cdot}3 \times 10^{-5}$%) and one larva per plankton sample (biomass percentages: 2·3–2·6 $\times 10^{-4}$%) respectively. Veliger larvae of the golden mussel (150·0 ± 20·0 μm) were recovered in all three replicates at one (biomass percentages: 4·7–5·9 $\times 10^{-4}$%) and five larva(e) per plankton sample (biomass percentages: 2·4–2·6 $\times 10^{-3}$%), while the Japanese sea cucumber (175·4 ± 159·0 μm) was detected in all replicates tested (biomass percentages: 0·23%–5·3 $\times 10^{-4}$%) except for one replicate at 0·01 larva per plankton sample (Table 1). Water lice (2020 × 528 μm) was recovered in all replicates at 0·1 (biomass percentages: 0·16%–0·18%) and 0·05 individual per plankton sample (biomass percentages: 6·7–9·1 $\times 10^{-2}$%). For the other two densities (0·01 and 0·001 individual per sample), it was detected in only one replicate (Table 1).

Discussion

SENSITIVITY OF 454 PYROSEQUENCING FOR DETECTION OF RARE SPECIES

Recent empirical studies have demonstrated that 454 pyrosequencing is a powerful discovery tool for the 'rare biosphere' (e.g. Fonseca *et al.* 2010; Behnke *et al.* 2011; Thomsen *et al.* 2011; Yu *et al.* 2012). However, the sensitivity of this method has not been addressed in aquatic communities heretofore. In this study, we tested its sensitivity by spiking plankton samples with various densities of species known to not already be present in the sample (i.e. freshwater species were added to marine plankton and vice versa). Our results showed that pyrosequencing can recover spiked indicator species with biomass percentage as low as approximately $2{\cdot}3 \times 10^{-5}$% in artificially assembled plankton communities (Table 1). This high sensitivity was also supported by pyrosequencing results from a complex plankton sample (Fig. 3b). Although we used 80 μm nets for plankton sampling, some much smaller taxa including Chlorophyta and Bacillariophyta were also recovered (Fig. 3b). We do not know whether these species were captured in net samples, or were attached to the surface of and/or in the gut of larger species that were processed.

When the sensitivity of pyrosequencing was tested by spiking plankton samples with various densities of known indicator species, not all replicates in a given treatment yielded the same results. For example, the bay scallop was recovered from two of three replicates at a spiked density of one larva per plankton sample. This could have resulted from (i) incorrect manipulation of these small larvae (i.e. larva was not successfully added) and/or poor preservation, and (ii) random PCR failure due to a small amount of starting DNA template. However, the first explanation cannot be applied to the setups using <1 larva/juvenile, mainly because the same larvae/juveniles, which were lysed and then partitioned into different plankton samples, were successfully recovered from some of these setups (Table 1). The lower detection limit of analytical sensitivity is determined as the endpoint dilution at which 50% of the tested samples are positive with 95% confidence (World Organization for Animal Health (OIE) 2009). Unfortunately, we could not standardize our results using this criterion, not only because of the high expense associated with numerous replicates for pyrosequencing, but also because of possibly varied detection limit resulting from numerous factors, such as DNA isolation method, PCR primers, different developmental stages of species, complexity and composition of communities, and sequencing depth. However, the inconsistent results for the same species here at least suggest that we may be

approaching the detection limit for the tested species under the experimental conditions used. Moreover, we observed different detection sensitivity for different species, such as water lice vs. other the three smaller species (Table 1). This may be mainly due to different PCR amplification efficiency among these species. PCR amplification efficiency is highly associated with the degree of mismatches between PCR primers and templates (von Wintzingerode, Göbel & Stackebrandt 1997). Different mismatch degree among these species, such as number of mismatches and their positions in primers, may result in different PCR amplification, leading to different detection sensitivity. Although we did not use very deep sequencing in this study (i.e. an equivalent of 1/24 PicoTiter plate for each sample), the sensitivity of the method was far beyond our expectations and what might be recovered with the traditional approach of counting 300–500 individuals randomly obtained from a bulk sample in a counting chamber with a microscope (e.g. Lecroq *et al*. 2011). We expect that pyrosequencing would be even more sensitive when using deeper sequencing (at a correspondingly greater cost), although eventually diversity in all samples will asymptote with increased sequencing effort.

A key question in pyrosequencing studies is whether singletons are artefacts or informative for describing the 'rare biosphere'. Some studies have confirmed that singletons are informative in reflecting unique lineages in communities (e.g. Kauserud *et al*. 2012), whereas others have argued that singletons are mainly caused by sequencing artefacts and should therefore be eliminated from data sets during data processing (e.g. Tedersoo *et al*. 2010). Our study showed that we recovered spiked species as singletons in several setups with a high dilution level, such as 0·001 individual of water lice per plankton sample and one larva of bay scallop per plankton sample (Table 1). In addition, pyrosequencing of the plankton sample collected from Hamilton Harbor showed that some taxonomic groups were only recovered as singletons, such as Nematoda and Desmidiales (Appendix S1). Errors and artefacts may inflate the number of low-abundance reads; however, these factors cannot create new taxonomic groups. All these results suggest that some singletons are, in fact, informative and valuable. Proper management of low-abundance reads such as singletons and doubletons in pyrosequencing data sets is crucial for extracting the most accurate information about the real 'rare biosphere' in communities, although technical difficulties still remain for accurately sorting informative reads from errors and artefacts.

UNIVERSAL PRIMERS FOR PYROSEQUENCING OF COMPLEX COMMUNITIES

The first crucial step for biodiversity assessment using 454 pyrosequencing is the selection of proper genetic markers and their universal primers, which can be used to amplify and differentiate a range of species. Ideal candidate genes should have both conserved and variable regions, as the former can be used to design universal PCR primers across taxa while the latter allow discrimination over a wide range of taxonomic levels. Ribosomal RNA genes such as nSSU satisfy these criteria (Fig. 2) and have been widely used in PCR-based biodiversity studies (see review by Medlin & Kooistra 2010). Indeed, nSSU sequence data have become one of the most voluminous in public databases (see review by Medlin & Kooistra 2010), which facilitates species annotation by BLAST searches. Among variable regions of nSSU, the V4 region is hypervariable among species and employed for biodiversity assessment (e.g. Bråte *et al*. 2010; Cheung *et al*. 2010). Although we designed the primers based on three taxonomic groups (i.e. Crustacea, Mollusca, Tunicata), the universality tests in this study confirmed that the primer pair was powerful and universal enough to amplify a wide range of taxonomic groups, including animals, plants (algae), fungi, prokaryotes (Cyanobacteria) and protists (Fig. 3b, Appendix S1). All these results revealed that V4-nSSU and its universal primers designed here are sufficiently robust to recover plankton biodiversity in detail.

Conclusions

Despite the popularity of pyrosequencing for assessing biodiversity over the past several years, this newly emerged method still requires efficiency testing for the communities that it purports to represent. In this study, we demonstrate extreme sensitivity of pyrosequencing using rare species spiked into complex plankton samples, as well as the detection of microscopic larvae in plankton samples using a small-scale pyrosequencing effort. Although the method was characterized using plankton, we expect that it can be applied to other communities such as benthic communities, mainly owing to the universal nature of V4-nSSU primers for amplification of major invertebrate groups of Crustacea, Mollusca and Tunicata. Collectively, we propose that 454 pyrosequencing based on universal V4-nSSU primer pair is a robust tool for detection of rare native or alien species in complex communities.

Acknowledgments

We dedicate this manuscript in memory of our late colleague and co-author Dr. Martin Hulák. Great thanks to Drs. X. Zhou, S. Shokralla, A. Lévesque for their valuable comments on early versions of the manuscript, to C. Lewis for his technical advice, and to Drs. T. Therriault, S. A. Bailey, A. Locke, C. McKindsey, C. DiBacco and D. Duchini for plankton sample collections. This work was supported by Discovery grants from Natural Sciences and Engineering Research Council of Canada (NSERC) to DDH, MEC and HJM, by the Canadian Aquatic Invasive Species Network (CAISN), an Ontario Ministry of Research and Innovation Early Researcher Award to MEC, an NSERC Discovery Accelerator Supplement to HJM, and by the '100 Talents Program' of the Chinese Academy of Sciences to AZ.

References

Barrett, R.D.H. & Hebert, P.D.N. (2005) Identifying spiders through DNA barcodes. *Canadian Journal of Zoology*, **83**, 481–491.

Behnke, A., Engel, M., Christen, R., Nebel, M., Klein, R.R. & Stoeck, T. (2011) Depicting more accurate pictures of protistan community complexity using pyrosequencing of hypervariable SSU rRNA gene regions. *Environmental Microbiology*, **13**, 340–349.

Bråte, J., Logares, R., Berney, C., Ree, D.K., Klaveness, D., Jakobsen, K.S. & Shalchian-Tabrizi, K. (2010) Freshwater Perkinsea and marine-freshwater colonizations revealed by pyrosequencing and phylogeny of environmental rDNA. *The International Society for Microbial Ecology Journal*, **4**, 1144–1153.

Briski, E., Cristescu, M.E., Bailey, S.A. & MacIsaac, H.J. (2011) Use of DNA barcoding to detect invertebrate invasive species from diapausing eggs. *Biological Invasions*, **13**, 1325–1340.

Cheung, M.K., Au, C.H., Chu, K.H., Kwan, H.S. & Wong, C.K. (2010) Composition and genetic diversity of picoeukaryotes in subtropical coastal waters as revealed by 454 pyrosequencing. *The International Society for Microbial Ecology Journal*, **4**, 1053–1059.

Creer, S. (2010) Second-generation sequencing derived insights into the temporal biodiversity dynamics of freshwater protists. *Molecular Ecology*, **19**, 2829–2831.

Crooks, J.A. & Soulé, M.E (1999) Lag times in population explosions of invasive species: Cases and implications. *Invasive Species and Biodiversity Management*. (eds O.T. Sandlund, P. J. Schei & A. Viken). Kluwer Academic Publishers, Dordrecht, the Netherlands.

Darling, J.A. & Mahon, A.R. (2011) From molecules to management: Adopting DNA-based methods for monitoring biological invasions in aquatic environments. *Environmental Research*, **111**, 978–988.

Edgar, R.C., Haas, B.J., Clemente, J.C., Quince, C. & Knight, R. (2011) UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics*, **27**, 2194–2200.

Fonseca, V.G., Carvalho, G.R., Sung, W., Johnson, H.F., Power, D.M., Neill, S. P., Packer, M., Blaxter, M.L., Lambshead, P.J.D., Thomas, W.K. & Creer, S. (2010) Second-generation environmental sequencing unmasks marine metazoan biodiversity. *Nature Communications*, **1**, 98.

Gu, W. & Swihart, R.K. (2004) Absent or undetected? Effects of non-detection of species occurrence on wildlife-habitat models. *Biological Conservation*, **116**, 195–203.

Harvey, C.T., Qureshi, S.A. & MacIsaac, H.J. (2009) Detection of a colonizing, aquatic, non-indigenous species. *Diversity and Distributions*, **15**, 429–437.

Hebert, P.D.N., Penton, E.H., Burns, J.M., Janzen, D.H. & Hallwachs, W. (2004) Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 14812–14817.

Jerde, C. L., Mahon, A. R., Chadderton, W. L. & Lodge, D. M. (2011) "Sight-unseen" detection of rare aquatic species using environmental DNA. *Conservation Letters*, **4**, 150–157.

Kauserud, H., Kumar, S., Brysting, A.K., Norden, J. & Carlsen, T. (2012) High consistency between replicate 454 pyrosequencing analyses of ectomycorrhizal plant root samples. *Mycorrhiza*, **22**, 309–315.

Kumar, S., Carlsen, T., Mevik, B., Enger, P., Blaalid, R., Shalchian-Tabrizi, K. & Kauserud, H. (2011) CLOTU: An online pipeline for processing and clustering of 454 amplicon reads into OTUs followed by taxonomic annotation. *BMC Bioinformatics*, **12**, 182.

Lecroq, B., Lejzerowicz, F., Bachar, D., Christen, R., Esling, P., Baerlocher, L., Østerås, M., Farinelli, L. & Pawlowski, J. (2011) Ultra-deep sequencing of foraminiferal microbarcodes unveils hidden richness of early monothalamous lineages in deep-sea sediments. *Proceedings of the National Academy of Sciences of the United States of America*, **108**, 13177–13182.

Li, W. & Godzik, A. (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, **22**, 1658–1659.

Lodge, D.M., Williams, S., MacIsaac, H.J., Hayes, K.R., Leung, B., Reichard, S., Mack, R.N., Moyle, P.B., Smith, M., Andow, D.A., Carlton, J.T. & McMichael, A. (2006) Biological invasions: recommendations for U.S. policy and management. *Ecological Applications*, **16**, 2035–2054.

Lodge, D.M., Turner, C.R., Jerde, C.L., Barnes, M.A., Chadderton, L., Egan, S. P., Feder, J.L., Mahon, A.R. & Pfrender, M.E. (2012) Conservation in a cup of water: estimating biodiversity and population abundance from environmental DNA. *Molecular Ecology*, **21**, 2555–2558.

Mackie, J.A. & Geller, J. (2010) Experimental parameters affecting quantitative PCR of *Artemia franciscana*: a model for a marine zooplanktonic target in natural plankton samples. *Limnology and Oceanography: Methods*, **8**, 337–347.

Magnuson, J.J., Benson, B.J. & McLain, A.S. (1994) Insights on species richness and turnover from long-term ecological research: fishes in north temperate lakes. *American Zoologist*, **34**, 437–451.

Mahon, A.R., Barnes, M.A., Senapati, S., Feder, J.L., Darling, J.A., Chang, H. & Lodge, D.M. (2011) Molecular detection of invasive species in heterogeneous mixtures using a microfluidic carbon nanotube platform. *PLoS ONE*, **6**, e17280.

McDonald, L.L. (2004) Sampling rare populations. In *Sampling rare or elusive species*. (ed W.L. Thompson), Pp. 11–42. Island Press, New York.

Medlin, L.K. & Kooistra, W.H.C.F. (2010) Methods to estimate the diversity in the marine photosynthetic protist community with illustrations from case studies: A review. *Diversity*, **2**, 973–014.

Parameswaran, P., Jalili, R., Tao, L., Shokralla, S., Gharizadeh, B., Ronaghi, M. & Fire2, A.Z. (2007) A pyrosequencing-tailored nucleotide barcode design unveils opportunities for large-scale sample multiplexing. *Nucleic Acids Research*, **35**, e130.

Pereira, H.M., Navarro, L.M. & Martins, I.S. (2012) Global Biodiversity Change: The Bad, the Good, and the Unknown. *Annual Review of Environment and Resources*, **37**, 25–50.

Pyšek, P. & Richardson, D.M. (2010) Invasive species, environmental change and management, and health. *Annual Review of Environment and Resources*, **35**, 25–55.

Ricciardi, A. (2007) Are modern biological invasions an unprecedented form of global change? *Conservation Biology*, **21**, 329–336.

Shade, A., Hogan, C.S., Klimowicz, A.K., Linske, M., McManus, P.S. & Handelsman, J. (2012) Culturing captures members of the soil rare biosphere. *Environmental Microbiology*. **14**, 2247–2252.

Shokralla, S., Zhou, X., Janzen, D.H., Hallwachs, W., Landry, J., Jacobus, L.M. & Hajibabaei, M. (2011) Pyrosequencing for mini-barcoding of fresh and old museum specimens. *PLoS ONE*, **6**, e21252.

Sogin, M.L., Morrison, H.G., Huber, J.A., Welch, D.M., Huse, S.M., Neal, P. R., Arrieta, J.M. & Herndl, G.J. (2006) Microbial diversity in the deep sea and the underexplored "rare biosphere". *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 12115–12120.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. & Kumar, S. (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution*, **28**, 2731–2739.

Tedersoo, L., Nilsson, R.H., Abarenkov, K., Jairus, T., Sadam, A., Saar, I., Bahram, M., Bechem, E., Chuyong, G. & Kõljalg, U. (2010) 454 Pyrosequencing and Sanger sequencing of tropical mycorrhizal fungi provide similar results but reveal substantial methodological biases. *New Phytologist*, **188**, 291–301.

Thomsen, P.F., Kielgast, J., Iversen, L.L., Wiuf, C., Rasmussen, M., Gilbert, M. T.P., Orlando, L. & Willerslev, E. (2011) Monitoring endangered freshwater biodiversity using environmental DNA. *Molecular Ecology*, **21**, 2565–2573.

Wilson, H.B., Joseph, L.N., Moore, A.L. & Possingham, H.P. (2011) When should we save the most endangered species? *Ecology Letters*, **14**, 886–890.

von Wintzingerode, F., Göbel, U.B. & Stackebrandt, E. (1997) Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis. *FEMS Microbiology Reviews*, **21**, 213–229.

World Organization for Animal Health (OIE) (2009) Principles and methods of validation of diagnostic assays for infectious diseases. *Manual of Diagnostic Tests for Aquatic Animals 2009*, pp. 10–30. OIE, Paris, France.

Yatsunenko, T., Rey, F.E., Manary, M.J., Trehan, I., Dominguez-Bello, M.G., Contreras, M., Magris, M., Hidalgo, G., Baldassano, R.N., Anokhin, A.P., Heath, A.C., Warner, B., Reeder, J., Kuczynski, J., Caporaso, J.G., Lozupone, C.A., Lauber, C., Clemente, J.C., Knights, D., Knight, R. & Gordon, J.I. (2012) Human gut microbiome viewed across age and geography. *Nature*, **486**, 222–227.

Yu, D.W., Ji, Y., Emerson, B.C., Wang, X., Ye, C., Yang, C. & Ding, Z. (2012) Biodiversity soup: metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring. *Methods in Ecology and Evolution*, **3**, 613–623.

Zhan, A., Bao, Z., Hu, X., Lu, W., Wang, S., Peng, W., Wang, M., Hui, M. & Hu, J. (2008) Accurate methods of DNA extraction and PCR-based genotyping for single scallop embryos/larvae long-preserved in ethanol. *Molecular Ecology Resources*, **8**, 790–795.

Received 20 October 2012; accepted 22 January 2013
Handling Editor: Sergei Kosakovsky Pond

Supporting Information

Additional Supporting Information may be found in the online version of this article.

Appendix S1. Statistics of sequence reads at 5% genetic divergence threshold and BLAST results for pyrosequencing of the plankton sample collected from Hamilton Harbor. For each similarity-based cluster, ID of the representative sequence, number of sequence reads, the most similar match in GenBank and taxonomic group determination based on BLAST, and representative sequence for each cluster are shown.

